The Econometric Modelling of Financial Time Series

Second edition

Terence C. Mills

Professor of Economics,
Loughborough University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1999

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1993
Reprinted 1994, 1996, 1997
Second edition 1999

Printed and bound in Great Britain at the University Press, Cambridge

Typeset in Times 10/12 [KW]

British Library Cataloguing in Publication Data
Mills, Terence C.
The econometric modelling of financial time series / Terence C.
Mills. – 2nd edn
p. cm.
Includes bibliographical references (p.).
ISBN 0-521-62413-4. – ISBN 0-521-62492-4 (pbk.)
1. Finance–Econometric models. 2. Time-series analysis.
3. Stochastic processes. I. Title
HG174.M55 1999
332′.01′5195–dc21 98-53587 CIP

First edition
ISBN 0521 41046 7 hardback
ISBN 0521 42257 4 paperback

Second edition
ISBN 0521 62413 4 hardback
ISBN 0521 62492 4 paperback

Contents

Preface to second edition *page* vii

1 Introduction 1

2 Univariate linear stochastic models: basic concepts 8
2.1 Stochastic processes, ergodicity and stationarity 8
2.2 Stochastic difference equations 11
2.3 ARMA processes 13
2.4 Linear stochastic processes 28
2.5 ARMA model building 28
2.6 Non-stationary processes and ARIMA models 37
2.7 ARIMA modelling 48
2.8 Forecasting using ARIMA models 53

3 Univariate linear stochastic models: further topics 61
3.1 Determining the order of integration of a time series 62
3.2 Decomposing time series: unobserved component models and signal extraction 99
3.3 Measures of persistence and trend reversion 107
3.4 Fractional integration and long memory processes 114

4 Univariate non-linear stochastic models 122
4.1 Martingales, random walks and non-linearity 122
4.2 Testing the random walk hypothesis 124
4.3 Stochastic volatility 126
4.4 ARCH processes 131
4.5 Other non-linear univariate models 153
4.6 Testing for non-linearity 171

5 **Modelling return distributions** 177
5.1 Descriptive analysis of three return series 177
5.2 Two models for return distributions 178
5.3 Determining the tail shape of a returns distribution 184
5.4 Empirical evidence on tail indices 188
5.5 Testing for covariance stationarity 193
5.6 Modelling the central part of returns distributions 196
5.7 Data analytic modelling of skewness and kurtosis 198
5.8 Distributional properties of absolute returns 200
5.9 Summary and further extensions 203

6 **Regression techniques for non-integrated financial time series** 205
6.1 Regression models 205
6.2 ARCH-in-mean regression models 218
6.3 Misspecification testing 221
6.4 Robust estimation 233
6.5 The multivariate linear regression model 235
6.6 Vector autoregressions 238
6.7 Variance decompositions, innovation accounting and structural VARs 245
6.8 Vector ARMA models 248

7 **Regression techniques for integrated financial time series** 253
7.1 Spurious regression 253
7.2 Cointegrated processes 260
7.3 Testing for cointegration in regression 268
7.4 Estimating cointegrating regressions 273
7.5 VARs with integrated variables 279
7.6 Causality testing in VECMs 297
7.7 Fully modified VAR estimation 299
7.8 Impulse response asymptotics in non-stationary VARs 302

8 **Further topics in the analysis of integrated financial time series** 306
8.1 Testing for a single long-run relationship 306
8.2 Common trends and cycles 310
8.3 Estimating permanent and transitory components of a VECM 315
8.4 Present value models, excess volatility and cointegration 318
8.5 Generalisations and extensions of cointegration and error correction models 334

Data appendix 340
References 342

1 Introduction

The aim of this book is to provide the researcher in financial markets with the techniques necessary to undertake the empirical analysis of financial time series. To accomplish this aim we introduce and develop both univariate modelling techniques and multivariate methods, including those regression techniques for time series that seem to be particularly relevant to the finance area.

Why do we concentrate exclusively on time series techniques when, for example, cross-sectional modelling plays an important role in empirical investigations of the Capital Asset Pricing Model (CAPM): see, as an early and influential example, Fama and MacBeth (1973)? Our answer is that, apart from the usual considerations of personal expertise and interest, plus manuscript length considerations, it is because time series analysis, in both its theoretical and empirical aspects, has been for many years an integral part of the study of financial markets, with empirical research beginning with the papers by Working (1934), Cowles (1933, 1944) and Cowles and Jones (1937).

Working focused attention on a previously noted characteristic of commodity and stock prices: namely, that they resemble cumulations of purely random changes. Cowles investigated the ability of market analysts and financial services to predict future price changes, finding that there was little evidence that they could. Cowles and Jones reported evidence of positive correlation between successive price changes but, as Cowles (1960) was later to remark, this was probably due to their taking monthly averages of daily or weekly prices before computing changes: a 'spurious correlation' phenomenon analysed by Working (1960).

The predictability of price changes has since become a major theme of financial research but, surprisingly, little more was published until Kendall's (1953) study, in which he found that the weekly changes in a wide variety of financial prices could not be predicted from either past changes in the series or from past changes in other price series. This seems to have been the first explicit reporting of this oft-quoted property of financial prices, although further impetus to research on price predictability was only provided by the publication of the papers by Roberts (1959) and Osborne (1959). The former presents a largely heuristic argument for why successive price changes should be independent, while the latter develops the proposition that it is not absolute price changes but the logarithmic price changes which are independent of each other: with the auxiliary assumption that the changes themselves are normally distributed, this implies that prices are generated as Brownian motion.

The stimulation provided by these papers was such that numerous articles appeared over the next few years investigating the hypothesis that price changes (or logarithmic price changes) are independent, a hypothesis that came to be termed the random walk model, in recognition of the similarity of the evolution of a price series to the random stagger of a drunk. Indeed, the term ‘random walk’ is believed to have first been used in an exchange of correspondence appearing in *Nature* in 1905 (see Pearson and Rayleigh, 1905), which was concerned about the optimal search strategy for finding a drunk who had been left in the middle of a field. The solution is to start exactly where the drunk had been placed, as that point is an unbiased estimate of the drunk’s future position since he will presumably stagger along in an unpredictable and random fashion.

The most natural way to state formally the random walk model is as

$$P_t = P_{t-1} + a_t \tag{1.1}$$

where P_t is the price observed at the beginning of time t and a_t is an error term which has zero mean and whose values are independent of each other. The price change, $\Delta P_t = P_t - P_{t-1}$, is thus simply a_t and hence is independent of past price changes. Note that, by successive backward substitution in (1.1), we can write the current price as the cumulation of all past errors, i.e.

$$P_t = \sum\nolimits_{i=1}^{t} a_i$$

so that the random walk model implies that prices are indeed generated by Working’s ‘cumulation of purely random changes’. Osborne’s model of Brownian motion implies that equation (1.1) holds for the logarithms of P_t and, further, that a_t is drawn from a zero mean normal distribution having constant variance.

Most of the early papers in this area are contained in the collection of Cootner (1964), while Granger and Morgenstern (1970) provide a detailed development and empirical examination of the random walk model and various of its refinements. Amazingly, much of this work had been anticipated by the French mathematician Louis Bachelier (1900, English translation in Cootner, 1964) in a remarkable Ph.D. thesis in which he developed an elaborate mathematical theory of speculative prices, which he then tested on the pricing of French government bonds, finding that such prices were consistent with the random walk model. What made the thesis even more remarkable was that it also developed many of the mathematical properties of Brownian motion which had been thought to have first been derived some years later in the physical sciences, particularly by Einstein! Yet, as Mandelbrot (1989) remarks, Bachelier had great difficulty in even getting himself a university appointment, let alone getting his theories disseminated throughout the academic community!

It should be emphasised that the random walk model is only a hypothesis about how financial prices move. One way in which it can be tested is by examining the autocorrelation properties of price changes: see, for example, Fama (1965). A more general perspective is to view (1.1) as a particular model within the class of autoregressive-integrated-moving average (ARIMA) models popularised by Box and Jenkins (1976). Chapter 2 thus develops the theory of such models within the general context of (univariate) linear stochastic processes.

We should avoid giving the impression that the only financial time series of interest are stock prices. There are financial markets other than those for stocks, most notably for bonds and foreign currency, but there also exist the various futures, commodity and derivative markets, all of which provide interesting and important series to analyse. For certain of these, it is by no means implausible that models other than the random walk may be appropriate or, indeed, models from a class other than the ARIMA. Chapter 3 discusses various topics in the general analysis of linear stochastic models, most notably that of determining the order of integration of a series and, associated with this, the appropriate way of modelling trends and structural breaks. It also considers methods of decomposing an observed series into two or more unobserved components and of determining the extent of the ‘memory’ of a series, by which is meant the behaviour of the series at low frequencies or, equivalently, in the very long run. A variety of examples taken from the financial literature are provided throughout the chapter.

During the 1960s much research was also carried out on the theoretical foundations of financial markets, leading to the development of the theory of efficient capital markets. As LeRoy (1989) discusses, this led to some serious questions being raised about the random walk hypothesis as a *theoretical* model of financial markets, the resolution of which required situating the hypothesis within a framework of economic equilibrium. Unfortunately, the assumption in (1.1) that price changes are independent was found to be too restrictive to be generated within a reasonably broad class of optimising models. A model that is appropriate, however, can be derived for stock prices in the following way (similar models can be derived for other sorts of financial prices, although the justification is sometimes different: see LeRoy, 1982). The return on a stock from t to $t+1$ is defined as the sum of the dividend yield and the capital gain, i.e., as

$$r_{t+1} = \frac{P_{t+1} + D_t - P_t}{P_t} \tag{1.2}$$

where D_t is the dividend paid during period t. Let us suppose that the expected return is constant, $E_t(r_{t+1}) = r$, where $E_t(\ )$ is the expectation conditional on information available at t: r_t is then said to be a *fair game*. Taking expectations at t of both sides of (1.2) and rearranging yields

$$P_t = (1+r)^{-1} E_t(P_{t+1} + D_t) \tag{1.3}$$

which says that the stock price at the beginning of period t equals the sum of the expected future price and dividend, discounted back at the rate r. Now assume that there is a mutual fund that holds the stock in question and that it reinvests dividends in future share purchases. Suppose that it holds h_t shares at the beginning of period t, so that the value of the fund is $x_t = h_t P_t$. The assumption that the fund ploughs back its dividend income implies that h_{t+1} satisfies

$$h_{t+1} P_{t+1} = h_t(P_{t+1} + D_t)$$

Thus

$$\begin{aligned} E_t(x_{t+1}) &= E_t(h_{t+1}P_{t+1}) = h_t E_t(P_{t+1} + D_t) \\ &= (1+r)h_t P_t = (1+r)x_t \end{aligned}$$

i.e., that x_t is a *martingale* (if, as is common, $r > 0$ we have that $E_t(x_{t+1}) \geq x_t$, so that x_t is a *submartingale*: LeRoy (1989, pp. 1593–4), however, offers an example in which r could be negative, in which case x_t

will be a *supermartingale*). LeRoy (1989) emphasises that price itself, without dividends added in, is not generally a martingale, since from (1.3) we have

$$r = E_t(D_t)/P_t + E_t(P_{t+1})/P_t - 1$$

so that only if the expected dividend–price ratio (or dividend yield) is constant, say $E_t(D_t)/P_t = d$, can we write P_t as the submartingale (assuming $r > d$)

$$E_t(P_{t+1}) = (1 + r - d)P_t$$

The assumption that a stochastic process, y_t say, follows a random walk is more restrictive than the requirement that y_t follows a martingale. The martingale rules out any dependence of the conditional expectation of Δy_{t+1} on the information available at t, whereas the random walk rules out not only this but also dependence involving the higher conditional moments of Δy_{t+1}. The importance of this distinction is thus evident: financial series are known to go through protracted quiet periods and also protracted periods of turbulence. This type of behaviour could be modelled by a process in which successive conditional variances of Δy_{t+1} (but *not* successive levels) are positively autocorrelated. Such a specification would be consistent with a martingale, but not with the more restrictive random walk.

Martingale processes are discussed in chapter 4, and lead naturally on to non-linear stochastic processes that are capable of modelling higher conditional moments, such as the autoregressive conditionally heteroskedastic (ARCH) model introduced by Engle (1982), stochastic variance models, and the bilinear process analysed by Granger and Andersen (1978). Also discussed in this chapter are a variety of other non-linear models, including Markov switching processes, smooth transitions and chaotic models, and the various tests of non-linearity that have been developed. The chapter also includes a discussion of the computer intensive technique of artificial neural network modelling. The various techniques are illustrated using exchange rates and stock price series.

The focus of chapter 5 is on the unconditional distributions of asset returns. The most noticeable feature of such distributions is their leptokurtic property: they have fat tails and high peakedness compared to a normal distribution. Although ARCH processes can model such features, much attention in the finance literature since Mandelbrot's (1963a, 1963b) path-breaking papers has concentrated on the possibility that returns are generated by a stable process, which has the property of having an infinite variance. Recent developments in statistical analysis have allowed a much deeper investigation of the tail shapes of empirical distributions, and methods of estimating tail shape indices are introduced and applied to a variety of returns series. The chapter then looks at the implications of fat-tailed distributions for testing the covariance stationarity assumption of time series analysis, data analytic methods of modelling skewness and kurtosis, and the impact of analysing transformations of returns, rather than the returns themselves.

The remaining three chapters focus on multivariate techniques of time series analysis, including regression methods. Chapter 6 concentrates on analysing the relationships between a set of *stationary* or, more precisely, *non-integrated*, financial time series and considers such topics as general dynamic regression, robust estimation, generalised methods of moments estimation, multivariate regression, vector autoregressions, Granger-causality, variance decompositions and impulse response analysis. These topics are illustrated with a variety of examples drawn from the finance literature: using forward exchange rates as optimal predictors of future spot rates, modelling the volatility of stock returns and the risk premium in the foreign exchange market, testing the CAPM, and investigating the interaction of the equity and gilt markets in the UK.

Chapter 7 concentrates on the modelling of *integrated* financial time series, beginning with a discussion of the spurious regression problem, introducing cointegrated processes and demonstrating how to test for cointegration, and then moving on to consider how such processes can be estimated. Vector error correction models are analysed in detail, along with associated issues in causality testing and impulse response analysis. The techniques introduced in this chapter are illustrated with extended examples analysing the market model and the interactions of the UK financial markets.

Finally, chapter 8 considers further issues in multivariate time series analysis, such as alternative approaches to testing for the presence of a long-run relationship, the analysis of both common cycles and trends, methods of computing permanent and transitory decompositions in a multivariate framework, and extensions to deal with infinite variance errors and structural breaks. Modelling issues explicit to finance are also discussed. Samuelson (1965, 1973) and Mandelbrot (1966) analysed the implications of equation (1.3), that the stock price at the beginning of time t equals the discounted sum of next period's expected future price and dividend, to show that this stock price equals the expected discounted, or present, value of *all* future dividends, i.e., that

$$P_t = \sum\nolimits_{i=0}^{\infty} (1 + r)^{-(i+1)} E_t\big(D_{t+i}\big) \tag{1.4}$$

which is obtained by recursively solving (1.3) forwards and assuming that $(1 + r)^{-n} E_t\big(P_{t+n}\big)$ converges to zero as $n \rightarrow \infty$. Present value models of the type (1.4) are analysed extensively in chapter 8, with the theme of whether stock markets are excessively volatile, perhaps containing speculative bubbles, being used extensively throughout the discussion and in a succession of examples, although the testing of the expectations hypothesis of the term structure of interest rates is also used as an example of the general present value framework.

Having emphasised earlier in this chapter that the book is exclusively about modelling financial time series, we should state at this juncture what the book is not about. It is certainly not a text on financial market theory, and any such theory is only discussed when it is necessary as a motivation for a particular technique or example. There are numerous texts on the theory of finance and the reader is referred to these for the requisite financial theory: two notable texts that contain both theory and empirical techniques are Campbell, Lo and MacKinlay (1997) and Cuthbertson (1996). Neither is it a textbook on econometrics. We assume that the reader already has a working knowledge of probability, statistics and econometric theory. Nevertheless, it is also non-rigorous, being at a level roughly similar to my *Time Series Techniques for Economists* (1990), in which references to the formal treatment of the theory of time series are provided.

When the data used in the examples throughout the book have already been published, references are given. Previous unpublished data are defined in the data appendix, which contains details on how they may be accessed. All standard regression computations were carried out using *EVIEWS 2.0* (EViews, 1995) or *MICROFIT* 4.0 (Pesaran and Pesaran, 1997). *PcFiml* 9.0 (Doornik and Hendry, 1997), *STAMP* 5.0 (Koopman *et al.*, 1995), *SHAZAM* (Shazam, 1993) and *COINT 2.0a* (Ouliaris and Phillips, 1995) were also used for particular examples and 'non-standard' computations were made using algorithms written by the author in *GAUSS 3.1.*